Fairness in Examination Timetabling: Student Preferences and Extended Formulations

A. Muklason[1], A.J. Parkes[1], B. McCollum[2], E. Özcan[1]

[1]University of Nottingham, [2]Queen's University of Belfast

{*abm, ajp, exo*}*@cs.nott.ac.uk*, *b.mccollum@qub.ac.uk*

August 26, 2014

PATAT 2014; York, United Kingdom, 26-29 August 2014

Outline

Introduction

- Well-known as challenging real world optimisation problems (NP-Hard).
- Could be considered as an assignment problems of exams into limited number resources (i.e. time-slots and rooms) subject to some restricted constrains.
- **Hard constraints**, e.g. Clash-free timetable (no student has to sit more than one exam in the same time). Determine the *feasibility* of solution (i.e. timetable).
- **Soft constraints**, e.g. spread out the timetable as evenly as possible. Determine the *quality* of solution.
- Standard examination timetabling problems only minimise average penalty per student that can lead Unfairness.

Research Question

Regarding examination timetable, what students consider to be best for them and also fair between students?

Related Works

- Important, since exams contribute significantly to students academic achievement.
- In our prior work [Muklason et. al., 2013] extended formulation was made in order to encourage fairness among the entire student body.
- [Mühlenthaler et. al., 2014] worked on investigating fairness in course timetabling problems.

Survey Results

General feedbacks:

- Very Happy (4%), Happy (37%), Moderate (47%), Unhappy (12%).
- 29 % believed that their examination timetable negatively affected their academic achievement.
- Reasons for not happy: two consecutive exams with less than 24 hours gap, the exams were not spread out evenly along exams period, had two hard exams scheduled too closed to each other.
- 61% believed that the examination timetable was fair amongst student and 10% said the opposite

Survey Results

Feedback that lack of attention in the state-of-the-art research in Exam timetabling:

- Students do have concerns about "fairness within a course", that is, fairness between students within their own course as opposed to only between students in the entire university.
- 43% strongly agreed and 41% agreed for "fairness within a course".
- 25% strongly agreed and 41% agreed for fairness amongst all student body.
- Students do consider some examinations harder than others (53% strongly agreed and 37% agreed) and would prefer a larger time gap before such hard examinations.

Extended Examination Timetabling Problem Formulation

In order to take findings from the survey into account, some extensions of ITC 2007 examination problem formulation are made:

- Introducing difficulty index for each exam representing how difficult the exam in three soft constraints formulation: Two Exams in a Row, Two Exams in a Day, and Period Spread.
- Promoting fairness as one of objective functions.

Extended Examination Timetabling Problem Formulation

Fairness Objective Function:
Jain Fairness Index (JFI) is adopted. Let C is a set of students, with their associated penalties: $P(C) = \{p_i\}$, σ_P^2: variance on the penalties, $\bar{P}$: mean value of penalties.'Relative Standard Deviation'(RSD): $RSD^2 = \sigma_P^2/\bar{P}^2$.

$$J(C) = \left(1 + RSD^2\right)^{-1} = \frac{\left(\sum_{i \in C} p_i\right)^2}{|C| \sum_{i \in C} p_i^2} \qquad (1)$$

The range of J(C) value is **1/**$|C|$ - **1**. J(C) should be maximised. Only penalty related to each individual student considered: **Two Exams in a Row, Two Exams in a Day, and Period Spread**.

Extended Examination Timetabling Problem Formulation

Fairness Within a Cohort Objective Function:
Let C_i is a set of students in cohort i of a total number k cohort, then the **candidate functions** could be:

$$(maximise) \sum_{i=1}^{k} w_i J(C_i) \tag{2}$$

Or, for some suitable value of p,

$$\text{(minimise)} \quad \sum_{i=1}^{k} w_i RSD^p(C_i) \tag{3}$$

The weight for each cohort, w_i could be introduced to show the different important level of fairness of differing cohort.

Extended Examination Timetabling Problem Formulation

Making trade-off between: Standard quality (Minimise soft constraints penalty), Overall Fairness, and "Fairness within a Cohort".

Which solution is preferable?

Soln	P1	P2	avg(P)	Jall	J1	J2	avg(J)
S1	{4,4}	{2,2}	**3**	0.9	1.0	1.0	**1.0**
S2	{4,2}	{4,2}	**3**	0.9	0.9	0.9	0.9

Table: Case 01

Extended Examination Timetabling Problem Formulation

Making trade-off between: Standard quality (Minimise soft constraints penalty), Overall Fairness, and "Fairness within a Cohort". **Which solution is preferable?**

Soln	P1	P2	avg(P)	Jall	J1	J2	avg(J)
S1	{8,8,9}	{2,2,2}	5.2	0.725	0.997	1.0	**0.998**
S2	{8,8,2}	{8,2,2}	**5.0**	0.735	0.82	0.67	0.742
S3	{7,7,9}	{4,3,3}	5.5	**0.852**	0.985	0.980	0.983

Table: Case 02

Experimental Methods

Overall, the methods consist of three phases:

- Generating initial feasible solution.
- Improving initial feasible solution quality.
- Enforcing fairness within the solution.

Phase I: Generating Initial Feasible Solution

Problem instances from Carter dataset were converted as special case problem in ITC 2007 dataset. Hence, both are solved with the same method.

- An initial feasible solution is constructed by using adaptive heuristic ordering (AHO).
- Basic Principle: The most 'difficult' (in term of the availability of feasible time-slots and room) exam is assigned to time-slot and room first.
- AHO approaches based on squeaky wheel optimisation (AHOSWO)[Abdulrahman et. al., 2014] was implemented.

Phase II: Improving The Initial feasible Solution Quality

- Selection hyper-heuristic approach implemented within Hyflex framework.
- Implemented mostly used low-level heuristics (LLH) for exam timetabling problem in the literature.
- Selection LLH strategy: **Reinforcement Learning**[Kaelbling et. al., 1996].
- Move acceptance strategy: **Great Deluge Algorithm** [Dueck, 1993]

Enforcing Fairness Stage

- Make maximisation of overall Jain Fairness Index as objective function to promote fairness amongst students body of University.
- Make maximisation average course Jain Fairness Index as objective function to promote fairness amongst students within the same cohort/course.

Experimental Results: Improving The Quality of Initial Solutions

The quality of solutions after quality Improvement phase: multiple Initial Solution Generation with AHOSWO + Iterated Hyper-heuristics with 13 Common LLH-RL-GD Algorithm (**Iterated AHO-RLGDHH**) vs benchmark:

	Iterated AHO-RLGDHH	Burke (2010)
CAR91	5.25	**5.03**
CAR92	4.53	**4.22**
EAR83	36.29	**36.06**
HEC92	12.23	**11.71**
KFU93	**15.67**	16.02
LSE91	11.41	**11.15**
PUR93	4.86	N.A
RYE92	10.8	**9.42**
STA83	160.52	**158.86**
TRE92	9	**8.37**
UTA92	3.58	**3.37**
UTE92	28.71	**27.99**
YOR83	**39.41**	39.53

Table: After Improvement Phase Toronto Benchmark Dataset

Note: The **Iterated AHO-RLGDHH** only outperforms Burke(2010) for 2 out 13 p. instances.

Experimental Results: Improving The Quality of Initial Solutions

The quality of solutions after quality Improvement phase: multiple Initial Solution Generation with AHOSWO + Iterated Hyper-heuristics with 13 Common LLH-RL-GD Algorithm (**Iterated AHO-RLGDHH**) vs benchmark:

Instance	Iterated AHO-RLGDHH	Muller(2007)
EXAM1	6920	**4370**
EXAM2	636	**400**
EXAM3	10712	**10049**
EXAM4	18423	**18141**
EXAM5	3522	**2988**
EXAM6	27740	**26585**
EXAM7	5198	**4213**
EXAM8	10141	**7742**
EXAM9	1428	**1030**
EXAM10	**15521**	16682
EXAM11	**33948**	34129
EXAM12	**5466**	5535

Table: After Improvement Phase for ITC 2007 Benchmark Dataset

Note: The **Iterated AHO-RLGDHH** only outperforms Muller(2007) for 3 out 12 p. instances.

Experimental Results: Promoting Fairness

The quality of solutions after enforcing fairness phase: Maximise over all JFI while preventing the solution from getting worse (in terms of standard quality):

Instance	Initial Solution Quality	Initial Solution JFI	Final Solution Quality	Final Solution JFI	Delta (%) Quality	Delta (%) JFI
CAR91	5.25	0.33	5.25	0.33	0.0	0.0
CAR92	4.53	0.29	4.53	0.29	0.0	0.0
EAR83	36.29	0.82	36.29	0.82	0.0	0.0
HEC92	12.23	0.53	12.23	0.53	0.0	0.0
KFU93	15.67	0.55	15.67	0.55	0.0	0.0
LSE91	11.41	0.52	11.41	0.52	0.0	0.0
PUR93	4.86	0.32	4.86	0.32	0.0	0.0
RYE92	10.8	0.4	**10.6**	0.4	-1.9	0.0
STA83	160.52	0.91	160.46	0.91	0.0	0.0
TRE92	9	0.44	9	0.44	0.0	0.0
UTA92	3.58	0.23	3.58	0.23	0.0	0.0
UTE92	28.71	0.78	**28.13**	**0.79**	-2.0	1.3
YOR83	39.41	0.74	39.37	0.74	-0.1	0.0
		Average			-0.3	0.1

Table: After Enforcing Fairness Phase Solution for Carter Dataset I (best solution from 21 runs)

Note: Preventing std. quality getting worse, the fairness of only 1 out 13 p. instances could be improved.

Experimental Results: Promoting Fairness

The quality of solutions after enforcing fairness phase: Maximise over all JFI while preventing the solution from getting worse (in terms of standard quality):

Instance	Initial Solution		Final Solution		Delta (%)	
	Quality	JFI	Quality	JFI	Quality	JFI
EXAM1	6920	0.44	**6838**	**0.55**	-1.2	25.0
EXAM2	636	0.06	636	**0.96**	0.0	1500.0
EXAM3	10712	0.09	10712	**0.1**	0.0	11.1
EXAM4	18423	0.35	18423	0.35	0.0	0.0
EXAM5	3522	0.23	3522	0.23	0.0	0.0
EXAM6	27740	0.25	27740	**0.28**	0.0	12.0
EXAM7	5198	0.2	5198	**0.47**	0.0	135.0
EXAM8	10141	0.44	10141	**0.77**	0.0	75.0
EXAM9	1428	0.43	**1427**	**0.87**	-0.1	102.3
EXAM10	15521	0.34	15521	**0.35**	0.0	2.9
EXAM11	33948	0.13	33948	**0.15**	0.0	15.4
EXAM12	5366	0.12	**5290**	0.12	-1.4	0.0
	AVG				-0.2	156.6

Table: After Enforcing Fairness Phase Solution for ITC 2007 I (best solution from 21 runs)

Note: Preventing std. quality getting worse, the fairness of 9 out 12 p. instances could be improved.

Experimental Results: Promoting Fairness

The quality of solutions after enforcing fairness phase: Maximise over all JFI while allowing the solution getting worse (in terms of standard quality):

Instance	Initial Solution		Final Solution		Delta (%)	
	Quality	JFI	Quality	JFI	Quality	JFI
CAR91	5.25	0.33	5.95	**0.35**	13	6
CAR92	4.53	0.29	5.19	**0.31**	15	7
EAR83	36.29	0.82	39.91	**0.84**	10	2
HEC92	12.23	0.53	13.83	**0.58**	13	9
KFU93	15.67	0.55	19.14	**0.67**	22	22
LSE91	11.41	0.52	12.71	**0.57**	11	10
PUR93	4.86	0.32	5	**0.33**	3	3
RYE92	10.8	0.4	11.49	**0.42**	6	5
STA83	160.52	0.91	167.46	**0.94**	4	3
TRE92	9	0.44	10.04	**0.47**	12	7
UTA92	3.58	0.23	3.98	**0.25**	11	9
UTE92	28.71	0.78	31.45	**0.81**	10	4
YOR83	39.41	0.74	44.37	**0.77**	13	4
		AVG			11	7

Table: After Enforcing Fairness Phase Solution for Carter Dataset II(Fairer solution causing the least increase in overall penalty from 21 runs)

Note: Allowing std. quality worse, the fairness of all 13 p. instances could be improved.

Experimental Results: Promoting Fairness

The quality of solutions after enforcing fairness phase: Maximise over all JFI while allowing the solution getting worse (in terms of standard quality):

Instance	Initial Solution		Final Solution		Delta (%)	
	Quality	JFI	Quality	JFI	Quality	JFI
EXAM1	6920	0.44	6900	**0.64**	-0.3	45.5
EXAM2	636	0.06	636	**0.95**	0.0	1483.3
EXAM3	10712	0.09	10742	**0.11**	0.3	22.2
EXAM4	18423	0.35	20203	**0.41**	9.7	17.1
EXAM5	3522	0.23	3578	**0.24**	1.6	4.3
EXAM6	27740	0.25	27765	**0.27**	0.1	8.0
EXAM7	5198	0.2	5198	**0.49**	0.0	145.0
EXAM8	10141	0.44	10172	**0.49**	0.3	11.4
EXAM9	1428	0.43	1433	**0.61**	0.4	41.9
EXAM10	15521	0.34	15764	**0.37**	1.6	8.8
EXAM11	33948	0.13	34308	**0.14**	1.1	7.7
EXAM12	5366	0.12	8914	**0.16**	66.1	33.3
		AVG			6.7	152.4

Table: After Enforcing Fairness Phase Solution for ITC 2007 II (Fairer solution causing the least increase in overall penalty from 21 runs)

Note: Allowing std. quality worse, the fairness of all 12 p. instances could be improved.

Experimental Results: Promoting Fairness within a Cohort

The quality of solutions after enforcing fairness within a cohort phase: Maximise average JFI per cohort while preventing the std. quality of solution getting worse:

Solution	Quality	JA	J1	J2	J3	AvgJC
Sol.0	**160.52**	**0.91**	**0.96**	**0.97**	**0.99**	**0.973**
Sol.1	159.11	0.91	0.96	0.98	0.99	0.977
Sol.2	160.46	0.91	0.96	0.98	1	0.980
Sol.3	**159.62**	**0.91**	**0.96**	**0.97**	**0.99**	**0.973**
Sol.4	165.31	0.93	0.96	0.99	1	0.983
Sol.5	167.46	0.94	0.97	0.99	0.99	0.983

Table: After Enforcing "Fairness Within a Cohort" *)

*) Problem instance: Adapted from an instance in Carter dataset, **STA83**. The 611 students from the original problem instance were randomly assigned into three different cohorts.

Note: 5 differing solutions from 42 runs with Sol.0 as an initial solution. Sol.0 and Sol.3 are dominated by Sol.1; in terms of quality, overall fairness, and "fairness within a cohort" the others are non-dominated solutions.

Conclusion

- The main contribution is to also account for 'fairness within a cohort of students', rather than only between the entire student body.
- Experimental results showed that all the solution could be made fairer by allowing the standard quality slightly worse.
- There is a trade-off between efficiency (core std. objective function) and fairness (including overall fairness and fairness within a cohort.

Future Works Plan

- Multi-objective Hyper-heuristics in order for producing pareto front between standard quality (Q) and fairness (F) both overall fairness and fairness within a cohort.
- Study which solutions of the multi-objective problem best match the student preferences, as well as the balance with requirements of the other stakeholders such as teachers and invigilators.

References

Leslie Pack Kaelbling, Michael L. Littman, and Andrew W. Moore. (1996)
Reinforcement learning: A survey.
*J. Artif. Int. Res.*4(1):237-285.

Carter, M. W., Laporte, G., & Lee, S. Y. (1996)
Examination timetabling: Algorithmic strategies and applications
*Journal of the Operational Research Society*47(3), 373-383.

McCollum, B., McMullan, P., Parkes, A. J., Burke, E. K., & Qu, R. (2012).
A new model for automated examination timetabling
*Ann Oper Res*194(1), 291-315.

Abdul Rahman, Syariza and Bargiela, Andrzej & Burke, Edmund K. & Ö zcan, Ender & McCollum, Barry & McMullan, Paul (2014).
Adaptive linear combination of heuristic orderings in constructing examination timetables
*European Journal of Operational Research*232(2), 287-297.

Mühlenthaler, M., Wanka, R. (2014).
Fairness in academic course timetabling
*Annals of Operations Research*2014, 1-18.

Muklason, A., Parkes, A.J., McCollum, B., Özcan, E. (2013).
Initial results on fairness in examination timetabling
In: Proc. of the 6th Multidisciplinary International Conference on Scheduling : Theory and Applications (MISTA) (2013).

Gunter Dueck. (1993).
New optimization heuristics: The great deluge algorithm and the record-to-record travel.
Journal of Computational Physics, 104(1),86-92.

The End

Thank You
Questions?